

RESEARCH ARTICLE

# A note on continuity and asymptotic consistency of measures of risk and variability

Niushan Gao and Foivos Xanthos 

Department of Mathematics, Toronto Metropolitan University, Toronto, M5B 2K3, Canada

**Corresponding author:** Foivos Xanthos; Email: [foivos@torontomu.ca](mailto:foivos@torontomu.ca)

**Received:** 1 September 2024; **Revised:** 28 October 2024; **Accepted:** 28 October 2024;

**First published online:** 16 December 2024

**Keywords:** Automatic continuity; strong consistency; risk measures; variability measures

**JEL codes:** 91G70; 91B30; 46E30

## Abstract

In this short note, we show that every convex, order-bounded above functional on a Fréchet lattice is automatically continuous. This improves a result in Ruszczyński and Shapiro ((2006) *Mathematics of Operations Research* **31**(3), 433–452.) and applies to many deviation and variability measures. We also show that an order-continuous, law-invariant functional on an Orlicz space is strongly consistent everywhere, extending a result in Krätschmer *et al.* ((2017) *Finance and Stochastics* **18**(2), 271–295.).

## 1. Automatic continuity

Since its introduction in the landmark paper (Artzner *et al.*, 1999), the axiomatic theory of risk measures has been a fruitful area of research. Among many topics, one particular direction is to investigate automatic continuity of risk measures. In general, automatic continuity has long been an interesting research topic in mathematics and probably originates from the fact that a real-valued convex function on an open interval is continuous. This well-known fact was later extended to the following theorem for real-valued convex functionals on general Banach lattices.

**Theorem** (Ruszczyński and Shapiro 2006). *A real-valued, convex, monotone functional on a Banach lattice is norm continuous.*

Recall that a functional  $\rho$  on a vector space  $\mathcal{X}$  is said to be *convex* if  $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$  for any  $X, Y \in \mathcal{X}$  and any  $\lambda \in [0, 1]$ . Recall also that a *vector lattice*  $\mathcal{X}$  is a real vector space with a partial order  $\leq$  such that (1) if  $X \leq Y$  in  $\mathcal{X}$ , then  $X + Z \leq Y + Z$  for any  $Z \in \mathcal{X}$  and  $\alpha X \leq \alpha Y$  for any  $\alpha \in \mathbb{R}$  with  $\alpha \geq 0$ ; (2) the supremum of any two elements exists in  $\mathcal{X}$ . We also denote  $X \leq Y$  by  $Y \geq X$ . In the case of a vector lattice of random variables, the order  $X \leq Y$  is always defined by  $X \leq Y$  a.s. See Aliprantis and Burkinshaw (2006) for standard terminology and facts regarding vector lattices. A *Banach lattice*  $\mathcal{X}$  is a vector lattice with a complete norm such that  $|X| \leq |Y|$  in  $\mathcal{X}$  implies  $\|X\| \leq \|Y\|$ . A functional  $\rho$  on a Banach lattice  $\mathcal{X}$  is said to be *increasing* if  $\rho(X) \leq \rho(Y)$  whenever  $X \leq Y$  in  $\mathcal{X}$ .  $\rho$  is said to be *decreasing* if  $-\rho$  is increasing.  $\rho$  is said to be *monotone* if it is either increasing or decreasing.

The above celebrated result of Ruszczyński and Shapiro has drawn extensive attention in optimization, operations research, and risk management. We refer the reader to Biagini and Frittelli (2009) for a version on Fréchet lattices and Farkas *et al.* (2014) for further literature on automatic norm continuity properties.

With law invariance, other types of continuity properties beyond norm continuity can be established. The theorem below is striking. Recall first that a functional  $\rho$  defined on a set  $\mathcal{X}$  of random variables

is said to be *law invariant* if  $\rho(X) = \rho(Y)$  whenever  $X, Y \in \mathcal{X}$  have the same distribution. Recall also that a functional  $\rho$  defined on a set  $\mathcal{X}$  of random variables is said to have the *Fatou property* if  $\rho(X) \leq \liminf_n \rho(X_n)$  whenever  $X_n \xrightarrow{o} X$  in  $\mathcal{X}$ . Here,  $X_n \xrightarrow{o} X$  in  $\mathcal{X}$ , termed as *order convergence* in  $\mathcal{X}$ ,<sup>1</sup> is used in the literature to denote dominated a.s. convergence in  $\mathcal{X}$ , that is,  $X_n \xrightarrow{a.s.} X$  and there exists  $Y \in \mathcal{X}$  such that  $|X_n| \leq Y$  a.s. for any  $n \in \mathbb{N}$ . The Fatou property is therefore just order lower semicontinuity.

**Theorem** (Jouini et al., 2006). *A real-valued, convex, monotone, law-invariant functional on  $L^\infty$  over a nonatomic probability space has the Fatou property. Consequently, it is  $(\sigma(L^\infty, L^1))$  lower semicontinuous and admits a dual representation via  $L^1$ .*

This theorem was recently extended by Chen et al. (2022) to general rearrangement-invariant spaces. See Chen et al. (2022), Theorems 2.2., 4.3, 4.7) for details. We also refer the reader to Shapiro (2013) for further interesting continuity properties of law-invariant risk measures.

In this section, we aim at extending the above theorem of Rusczyński and Shapiro on norm continuity of convex functionals. Specifically, we show that the monotonicity assumption can be significantly relaxed to the following notion on order boundedness.

**Definition 1.1.** *Let  $\mathcal{X}$  be a vector lattice. For  $U, V \in \mathcal{X}$  with  $U \leq V$ , the order interval  $[U, V]$  is defined by:*

$$[U, V] = \{X \in \mathcal{X} : U \leq X \leq V\}.$$

*A functional  $\rho: \mathcal{X} \rightarrow \mathbb{R}$  is said to be order bounded above if it is bounded above on all order intervals.*

Monotone functionals are easily seen to be order bounded above. While risk measures are usually assumed to be monotone, many important functionals used in finance, insurance, and other disciplines are not necessarily monotone.

**Example 1.2.** *General deviation measures were introduced in Rockafellar et al. (2006) as functionals  $\rho: L^2 \rightarrow [0, +\infty]$  satisfying subadditivity, positive homogeneity,  $\rho(X + c) = \rho(X)$  for every  $X \in L^2$  and  $c \in \mathbb{R}$ , and  $\rho(X) > 0$  for nonconstant  $X$ . They are usually not monotone but may be order bounded above. A specific example is standard deviation and semideviations. Recall that for a random variable  $X \in L^2$ , its standard deviation and upper and lower semideviations are given by:*

$$\sigma(X) = \|X - \mathbb{E}[X]\|_{L^2}, \quad \sigma_+(X) = \|(X - \mathbb{E}[X])^+\|_{L^2}, \quad \sigma_-(X) = \|(X - \mathbb{E}[X])^-\|_{L^2},$$

*respectively. They are well known to be convex. It is also easy to see that they are neither increasing nor decreasing. We show that they are all order bounded above on  $L^2$ . Indeed, take any order interval  $[U, V] \subset L^2$  and any  $X \in [U, V]$ . The desired order boundedness property is immediate by the following inequalities:*

$$\begin{aligned} U - \mathbb{E}[V] &\leq X - \mathbb{E}[X] \leq V - \mathbb{E}[U] \\ 0 &\leq (X - \mathbb{E}[X])^+ \leq (V - \mathbb{E}[U])^+ \\ 0 &\leq (X - \mathbb{E}[X])^- \leq (U - \mathbb{E}[V])^- \end{aligned}$$

**Example 1.3.** *General variability measures were introduced in Bellini et al. (2022) as law-invariant, positive-homogeneous functionals that vanish on constants. Many of them are also order-bounded above, although usually not monotone. In fact, all the three one-parameter families of variability measures in Bellini et al. (2022, Section 2.3) are easily seen to be order bounded above but not monotone. Let us discuss in details the class of inter-ES differences  $\Delta_p^{\text{ES}}$ ,  $p \in (0, 1)$ , on  $L^1$ . Let  $X \in L^1$  and  $p \in (0, 1)$ . Recall the right and left expected shortfalls of  $X$ :*

$$\text{ES}_p(X) = \frac{1}{1-p} \int_p^1 F_X^{-1}(t) \, dt, \quad \text{ES}_p^-(X) = \frac{1}{p} \int_0^p F_X^{-1}(t) \, dt,$$

<sup>1</sup>The term order convergence originates from the theory of vector lattices (see e.g., Aliprantis and Burkinshaw, 2006). We note here that in our definition, we do not require that  $\mathcal{X}$  is a vector lattice.

where  $F_X^{-1}(t) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq t\}$  is the left quantile function of  $X$ . The inter-ES difference  $\Delta_p^{\text{ES}}$  is defined by:

$$\Delta_p^{\text{ES}}(X) := \text{ES}_p(X) - \text{ES}_{1-p}^-(X) = \text{ES}_p(X) + \text{ES}_p(-X).$$

$\Delta_p^{\text{ES}}$  is clearly convex. Take any order interval  $[U, V] \subset L^1$  and any  $X \in [U, V]$ . By the monotonicity of expected shortfall, we get

$$\Delta_p^{\text{ES}}(X) \leq \text{ES}_p(V) + \text{ES}_p(-U) < \infty.$$

This proves that  $\text{ES}_p$  is order-bounded above on  $L^1$ . Now suppose that  $U \in L^1$  follows a uniform distribution on  $[-1, 1]$ . Then

$$U \leq 1, \quad \Delta_p^{\text{ES}}(U) = 2\text{ES}_p(U) > 0 = \Delta_p^{\text{ES}}(1),$$

$$U \geq -1, \quad \Delta_p^{\text{ES}}(U) = 2\text{ES}_p(U) > 0 = \Delta_p^{\text{ES}}(-1).$$

Thus,  $\Delta_p^{\text{ES}}$  is not monotone.

Our main result in this section is as follows. Recall first that a topological vector space  $(\mathcal{X}, \tau)$  is a *Fréchet lattice* if  $\mathcal{X}$  is a vector lattice and  $\tau$  is induced by a complete metric such that 0 has a fundamental system of convex solid neighborhoods. A neighborhood  $\mathcal{V}$  of 0 in  $\mathcal{X}$  is solid if  $Y \in \mathcal{V}$  whenever  $X \in \mathcal{V}$  and  $|Y| \leq |X|$ . All  $L^p$  spaces and Orlicz spaces equipped with their natural norm are Banach lattices and, in particular, are Fréchet lattices.

**Theorem 1.4.** *Let  $(\mathcal{X}, \tau)$  be a Fréchet lattice. Let  $\rho: \mathcal{X} \rightarrow \mathbb{R}$  be a convex, order-bounded above functional. Then  $\rho$  is  $\tau$ -continuous.*

Recall that monotone functionals are order bounded above and Banach lattices are Fréchet. Thus, Theorem 1.4 includes the preceding theorem of Ruzsčzyński and Shapiro as a special case.

*Proof of Theorem 1.4.* Let  $(X_n)$  and  $X$  be such that  $X_n \xrightarrow{\tau} X$  in  $\mathcal{X}$ . We want to show that  $\rho(X_n) \rightarrow \rho(X)$ . Suppose otherwise that  $\rho(X_n) \not\rightarrow \rho(X)$ . By passing to a subsequence, we may assume that

$$|\rho(X_n) - \rho(X)| > \varepsilon_0 \quad \text{for some } \varepsilon_0 > 0 \text{ and any } n \in \mathbb{N}. \quad (1.1)$$

Let  $(\mathcal{V}_n)$  be a basis of 0 for  $\tau$  consisting of solid neighborhoods such that  $\mathcal{V}_{n+1} + \mathcal{V}_{n+1} \subseteq \mathcal{V}_n$  for each  $n \in \mathbb{N}$ . By passing to a further subsequence of  $(X_n)$ , we may assume that  $n|X_n - X| \in \mathcal{V}_n$  for any  $n \geq 1$ . Put  $W_n = \sum_{i=1}^n i|X_i - X|$ . For any  $n, m \in \mathbb{N}$ , we have

$$W_{n+m} - W_n = \sum_{i=n+1}^{n+m} i|X_i - X| \in \mathcal{V}_{n+1} + \mathcal{V}_{n+2} + \cdots + \mathcal{V}_{n+m} \subseteq \mathcal{V}_n.$$

Thus,  $(W_n)$  is a  $\tau$  Cauchy sequence. By the completeness of  $\tau$ , there exists  $Y \in \mathcal{X}$  such that  $W_n \xrightarrow{\tau} Y$ . Since  $(W_n)$  is an increasing sequence,  $W_n \uparrow Y$  (Aliprantis and Burkinshaw, 2006, Theorem 3.46). In particular, it follows that  $W_n \leq Y$  so that

$$|X_n - X| \leq \frac{1}{n} Y \quad \text{for any } n \in \mathbb{N}. \quad (1.2)$$

Moreover, since  $\rho$  is order bounded above on  $[X - Y, X + Y] = X + [-Y, Y]$ , there exists a real number  $M > 0$  such that

$$\rho(X + Z) \leq M \quad \text{for any } Z \in [-Y, Y], \text{ that is, whenever } |Z| \leq Y. \quad (1.3)$$

Now fix any  $\varepsilon > 0$ . Put  $N = \lfloor \frac{1}{\varepsilon} \rfloor + 1$ . By (1.2),

$$\frac{1}{\varepsilon} |X_n - X| \leq Y \quad \text{for any } n \geq N. \quad (1.4)$$

On the one hand, by the convexity of  $\rho$  and the representation

$$X_n = (1 - \varepsilon)X + \varepsilon \left( X + \frac{1}{\varepsilon}(X_n - X) \right),$$

we have

$$\rho(X_n) \leq (1 - \varepsilon)\rho(X) + \varepsilon \rho \left( X + \frac{1}{\varepsilon}(X_n - X) \right),$$

implying that

$$\rho(X_n) - \rho(X) \leq \varepsilon \left( \rho \left( X + \frac{1}{\varepsilon}(X_n - X) \right) - \rho(X) \right).$$

This together with (1.4) and (1.3) implies that

$$\rho(X_n) - \rho(X) \leq \varepsilon(M - \rho(X)) \quad \text{for any } n \geq N. \quad (1.5)$$

On the other hand, by the convexity of  $\rho$  and

$$2X - X_n = (1 - \varepsilon)X + \varepsilon \left( X + \frac{1}{\varepsilon}(X - X_n) \right),$$

we have as before that

$$\begin{aligned} \rho(2X - X_n) &\leq (1 - \varepsilon)\rho(X) + \varepsilon \rho \left( X + \frac{1}{\varepsilon}(X - X_n) \right) \\ &\leq (1 - \varepsilon)\rho(X) + \varepsilon M, \end{aligned}$$

for any  $n \geq N$ . In particular,

$$\rho(2X - X_n) - \rho(X) \leq \varepsilon(M - \rho(X)) \quad \text{for any } n \geq N.$$

By  $X = \frac{1}{2}X_n + \frac{1}{2}(2X - X_n)$  and the convexity of  $\rho$ , we also get

$$\rho(X) \leq \frac{1}{2}\rho(X_n) + \frac{1}{2}\rho(2X - X_n)$$

so that

$$\rho(X) - \rho(X_n) \leq \rho(2X - X_n) - \rho(X).$$

It follows that

$$\rho(X) - \rho(X_n) \leq \varepsilon(M - \rho(X)) \quad \text{for any } n \geq N. \quad (1.6)$$

Combining (1.5) and (1.6), we have

$$|\rho(X) - \rho(X_n)| \leq \varepsilon(M - \rho(X)) \quad \text{for any } n \geq N.$$

Hence,  $\rho(X_n) \rightarrow \rho(X)$ . This contradicts (1.1) and completes the proof.  $\square$

## 2. Strong consistency

In this section, we discuss the strong consistency of estimating the risk  $\rho(X)$  using estimates drawn from the empirical distributions. This problem has been studied for general convex risk measures on  $L^p$  spaces and Orlicz hearts in Shapiro (2013) and Krätschmer *et al.* (2014), respectively. We are motivated to study it on general Orlicz spaces.

Throughout this section, fix a nonatomic probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $L^0$  be the space of all random variables on  $\Omega$ , with a.s. equal random variables identified as the same. Let  $\mathcal{X}$  be a subset of  $L^0$ . Denote the set of distributions of all random variables in  $\mathcal{X}$  by:

$$\mathcal{M}(\mathcal{X}) = \{\mathbb{P} \circ X^{-1} : X \in \mathcal{X}\}.$$

Recall that a law-invariant functional  $\rho$  on  $\mathcal{X}$  induces a natural mapping  $\mathcal{R}_\rho$  on  $\mathcal{M}(\mathcal{X})$  by:

$$\mathcal{R}_\rho(\mathbb{P} \circ X^{-1}) = \rho(X), \quad \text{for any } X \in \mathcal{X}.$$

Recall that a sequence  $(X_n)$  of random variables is said to be stationary if for any  $k, n \in \mathbb{N}$  and any  $x_1, \dots, x_n \in \mathbb{R}$ , it holds that  $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_{k+1} \leq x_1, \dots, X_{k+n} \leq x_n)$ . Let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra on  $\mathbb{R}^{\mathbb{N}}$ . A set  $A \in \mathcal{F}$  is said to be invariant if there exists  $B \in \mathcal{B}$  such that  $A = \{(X_n)_{n \geq k} \in B\}$  for every  $k \in \mathbb{N}$ . A stationary sequence is said to be ergodic if every invariant set has probability 0 or 1. Birkhoff's ergodic theorem states that the arithmetic averages of a stationary ergodic sequence  $(X_n)$  converge a.s. to  $\mathbb{E}[X_1]$  whenever  $\mathbb{E}[|X_1|] < \infty$ . See Breiman (1991, Section 6.7) for more facts regarding stationary and ergodic processes.

Let  $\mathcal{X}$  be a subset of  $L^0$  containing  $L^\infty$ . Take any  $X \in \mathcal{X}$ . Let  $(X_n)$  be a stationary ergodic sequence of random variables with the same distribution as  $X$ . We denote the *empirical distribution* of  $X$  arising from  $X_1, \dots, X_n$  by:

$$\widehat{m}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i};$$

here  $\delta_x$  is the Dirac measure on  $\mathbb{R}$  at  $x$ . Since  $L^\infty \subset \mathcal{X}$ ,  $\widehat{m}_n \in \mathcal{M}(L^\infty) \subset \mathcal{M}(\mathcal{X})$ . This allows us to consider the corresponding *empirical estimate* for  $\rho(X)$ :

$$\widehat{\rho}_n := \mathcal{R}_\rho(\widehat{m}_n);$$

We say that  $\rho$  is *strongly consistent* at  $X$  if for any stationary ergodic sequence of random variables with the same distribution as  $X$ :

$$\widehat{\rho}_n = \mathcal{R}_\rho(\widehat{m}_n) \xrightarrow{a.s.} \rho(X).$$

We refer to Shapiro (2013), Krättschmer *et al.* (2014, 2017), and the references therein for literature on strong (and weak) consistency of risk measures. In particular, we are interested in the following result.

**Theorem** (Krättschmer *et al.*, 2014, Theorem 2.6). *A norm-continuous, law-invariant functional on an Orlicz heart is strongly consistent everywhere.*

We remark that this theorem was originally stated for real-valued, law-invariant, convex risk measures. In their definition in Krättschmer *et al.* (2014), convex risk measures are assumed to be monotone and thus are norm continuous by the aforementioned theorem of Ruszczyński and Shapiro. A quick examination of the proof of Krättschmer *et al.* (2014, Theorem 2.6) shows that norm continuity and law invariance are the only ingredients of the functional used.

Let's recall the definitions of Orlicz spaces and hearts. A function  $\Phi: [0, \infty) \rightarrow [0, \infty)$  is called an *Orlicz function* if it is nonconstant, convex, increasing, and  $\Phi(0) = 0$ . The *Orlicz space*  $L^\Phi$  is the space of all  $X \in L^0$  such that the *Luxemburg norm* is finite:

$$\|X\|_\Phi := \inf \left\{ \frac{1}{\lambda} : \lambda > 0 \text{ and } \mathbb{E} [\Phi(\lambda|X|)] \leq 1 \right\} < \infty.$$

The *Orlicz heart*  $H^\Phi$  is a subspace of  $L^\Phi$  defined by:

$$H^\Phi := \{X \in L^0 : \mathbb{E} [\Phi(\lambda|X|)] < \infty \text{ for any } \lambda > 0\}.$$

We refer to Edgar and Sucheston (1992) for standard terminology and facts on Orlicz spaces. Risk measures on Orlicz spaces have been studied extensively; see, for example, Bellini *et al.* (2021), Bellini and Rosazza Gianin (2012), Biagini *et al.* (2011), Gao *et al.* (2018, 2019, 2020), Gao and Xanthos (2018), and the references therein.

The above theorem in conjunction with Theorem 1.4 immediately yields the following result, which improves Krättschmer *et al.*, (2014), Theorem 2.6.

**Corollary 2.1.** *A convex, law-invariant, order-bounded-above functional on an Orlicz heart is strongly consistent everywhere.*

The above theorem of Krättschmer *et al.* is essentially due to the fact that for any  $X \in H^\Phi$ , and for a.e.  $\omega \in \Omega$ , there exist a random variable  $X^\omega$  on  $\Omega$  with same distribution as  $X$  and a sequence of random variables  $(X_n^\omega)$  on  $\Omega$  with distributions  $\widehat{m}_n(\omega)$ 's such that

$$\|X_n^\omega - X^\omega\|_\Phi \rightarrow 0. \quad (2.1)$$

This, however, does not hold for arbitrary random variables in a general Orlicz space  $L^\Phi$ . Specifically, when  $\Phi$  fails the  $\Delta_2$ -condition, there exists  $X \in L^\Phi \setminus H^\Phi$ . For this  $X$ , (2.1) must fail:  $X_n^\omega$  takes only at most  $n$  values and thus is a simple random variable lying in  $H^\Phi$ ; therefore, (2.1) would imply  $X \in H^\Phi$  as well.

We extend the theorem of Krättschmer *et al.* as follows. Recall first that on a set  $\mathcal{X} \subset L^0$ , a functional  $\rho: \mathcal{X} \rightarrow \mathbb{R}$  is said to be *order continuous* at  $X \in \mathcal{X}$  if  $\rho(X_n) \rightarrow \rho(X)$  whenever  $X_n \xrightarrow{o} X$  in  $\mathcal{X}$ . In the literature, order continuity is also termed as the *Lebesgue property*.

**Theorem 2.2.** *An order-continuous, law-invariant functional on an Orlicz space is strongly consistent everywhere.*

For the proof of Theorem 2.2, we need to establish a few technical lemmas, which along the way also reveal why order continuity is the most natural condition for general Orlicz spaces. For an Orlicz function  $\Phi$ , the Young class is defined by:

$$Y^\Phi := \{X \in L^0: \mathbb{E}[\Phi(|X|)] < \infty\}.$$

It is easy to see that  $H^\Phi \subset Y^\Phi \subset L^\Phi$ . As in Krättschmer *et al.* (2014), we use the term  $\Phi$ -weak topology in place of the  $\Phi(|\cdot|)$ -weak topology on  $\mathcal{M}(Y^\Phi)$  for brevity. This topology is metrizable (see e.g., Föllmer and Schied 2011, Corollary A.45). For the special case where  $\Phi(x) = \frac{x^p}{p}$  for some  $1 \leq p < \infty$ , the  $\Phi$ -weak topology generates the Wasserstein metric of order  $p$  (see e.g., Villani 2021, Theorem 7.12). Moreover, for a sequence  $(\mu_n) \subset \mathcal{M}(Y^\Phi)$  and  $\mu_0 \in \mathcal{M}(Y^\Phi)$ ,  $(\mu_n)$  converges  $\Phi$ -weakly to  $\mu_0$ , written as  $\mu_n \xrightarrow{\Phi\text{-weakly}} \mu_0$ , iff

$$\mu_n \xrightarrow{\text{weakly}} \mu \text{ and } \int \Phi(|x|)\mu_n(dx) \rightarrow \int \Phi(|x|)\mu_0(dx).$$

The following Skorohod representation for  $\Phi$ -weak convergence is a general order version of Krättschmer *et al.* (2014, Theorem 3.5) and Krättschmer *et al.* (2017, Theorem 6.1) beyond the Orlicz heart and without any restrictions on  $\Phi$ .

**Lemma 2.3.**

- (i) Let  $(\mu_n)$  be a sequence in  $\mathcal{M}(Y^\Phi)$  that converges  $\Phi$ -weakly to some  $\mu_0 \in \mathcal{M}(Y^\Phi)$ . Then there exist a subsequence  $(\mu_{n_k})$  of  $(\mu_n)$ , a sequence  $(X_k)$  in  $Y^\Phi$  and  $X \in Y^\Phi$  such that  $X_k$  has distribution  $\mu_{n_k}$  for each  $k \in \mathbb{N}$ ,  $X$  has distribution  $\mu_0$ , and  $X_k \xrightarrow{o} X$  in  $Y^\Phi$ .
- (ii) Let  $(X_n)$  in  $Y^\Phi$  and  $X \in Y^\Phi$  be such that  $X_n \xrightarrow{o} X$  in  $Y^\Phi$ . Then  $\mu_n \xrightarrow{\Phi\text{-weakly}} \mu_0$ , where  $\mu_n$ 's are the distributions of  $X_n$ 's and  $\mu_0$  is the distribution of  $X$ , respectively.

*Proof.* We start with the following observation. Since  $\Phi$  is continuous and increasing, for any sequence  $(X_n)$  in  $Y^\Phi$ , we have

$$\mathbb{E}\left[\Phi\left(\sup_{n \in \mathbb{N}} |X_n|\right)\right] = \mathbb{E}\left[\sup_{n \in \mathbb{N}} \Phi(|X_n|)\right] \quad (2.2)$$

(i). Take  $(\mu_n)$  in  $\mathcal{M}(Y^\Phi)$  that converges  $\Phi$ -weakly to  $\mu_0 \in \mathcal{M}(Y^\Phi)$ . Since the probability space is nonatomic, the classical Skorohod representation yields  $(X_n) \subset Y^\Phi$  and  $X \in Y^\Phi$  such that  $X_n \sim \mu_n$  for every  $n \in \mathbb{N}$ ,  $X \sim \mu_0$ , and  $X_n \xrightarrow{a.s.} X$ . Clearly,

$$\mathbb{E}[\Phi(|X|)] = \int \Phi(|x|)\mu_0(dx) = \lim_n \int \Phi(|x|)\mu_n(dx) = \lim_n \mathbb{E}[\Phi(|X_n|)] < \infty. \quad (2.3)$$

Since  $\Phi$  is continuous, we also have that  $\Phi(|X_n|) \xrightarrow{a.s.} \Phi(|X|)$ . This combined with (2.3) yields (see Aliprantis and Burkinshaw, 2006, Theorem 31.7) that

$$\|\Phi(|X_n|) - \Phi(|X|)\|_{L^1} \rightarrow 0.$$

Passing to a subsequence we may assume that

$$\sum_{n=1}^{\infty} \|\Phi(|X_n|) - \Phi(|X|)\|_{L^1} < \infty$$

so that

$$\sum_{n=1}^{\infty} |\Phi(|X_n|) - \Phi(|X|)| \in L^1.$$

In particular,

$$\sup_{n \in \mathbb{N}} |\Phi(|X_n|) - \Phi(|X|)| \in L^1.$$

It follows from  $\Phi(|X_n|) \leq |\Phi(|X_n|) - \Phi(|X|)| + \Phi(|X|)$  that  $\sup_{n \in \mathbb{N}} \Phi(|X_n|) \in L^1$ . Hence, by (2.2),  $\mathbb{E}[\Phi(\sup_{n \in \mathbb{N}} |X_n|)] = \mathbb{E}[\sup_{n \in \mathbb{N}} \Phi(|X_n|)] < \infty$ . That is,  $\sup_{n \in \mathbb{N}} |X_n| \in Y^\Phi$ ; equivalently,  $(X_n)$  is dominated in  $Y^\Phi$ . In particular, we have  $X_n \xrightarrow{o} X$  in  $Y^\Phi$ .

(ii). Let  $(X_n)$  be such that  $X_n \xrightarrow{o} X$  in  $Y^\Phi$  and  $\mu_n$  be the distribution of  $X_n$  for each  $n$ ,  $\mu_0$  be the distribution of  $X$ . We clearly have  $\mu_n \xrightarrow{\text{weakly}} \mu$  and by the continuity of  $\Phi$ , we get  $\Phi(|X_n|) \xrightarrow{a.s.} \Phi(|X|)$ . Since  $(X_n)$  is dominated in  $Y^\Phi$ ,  $\sup_{n \in \mathbb{N}} |X_n| \in Y^\Phi$ . Thus in view of (2.2), we get

$$\mathbb{E} \left[ \sup_{n \in \mathbb{N}} \Phi(|X_n|) \right] = \mathbb{E} \left[ \Phi \left( \sup_{n \in \mathbb{N}} |X_n| \right) \right] < \infty,$$

that is,  $\sup_{n \in \mathbb{N}} \Phi(|X_n|) \in L^1$ . By the dominated convergence theorem, we get

$$\int \Phi(|x|) \mu_0(dx) = \mathbb{E}[\Phi(|X|)] = \lim_n \mathbb{E}[\Phi(|X_n|)] = \lim_n \int \Phi(|x|) \mu_n(dx).$$

This proves  $\mu_n \xrightarrow{\Phi\text{-weakly}} \mu_0$ . □

The lemma below reveals the essential and natural role of order continuity.

**Lemma 2.4.** *Let  $\rho: Y^\Phi \rightarrow \mathbb{R}$  be law invariant. The following are equivalent.*

- (i)  $\mathcal{R}_\rho$  is continuous on  $\mathcal{M}(Y^\Phi)$  with the  $\Phi$ -weak topology.
- (ii)  $\rho$  is order continuous on  $Y^\Phi$ .

*Proof.* (ii)  $\implies$  (i). Suppose that (ii) holds but (i) fails. Recall that the  $\Phi$ -weak topology is metrizable. Thus, we can find a sequence  $(\mu_n)$  and  $\mu_0$  in  $\mathcal{M}(Y^\Phi)$  such that  $\mu_n \xrightarrow{\Phi\text{-weakly}} \mu_0$  but  $\mathcal{R}_\rho(\mu_n) \not\rightarrow \mathcal{R}_\rho(\mu_0)$ . Passing to a subsequence, we may assume that

$$|\mathcal{R}_\rho(\mu_n) - \mathcal{R}_\rho(\mu_0)| \geq \varepsilon_0, \tag{2.4}$$

for some  $\varepsilon_0 > 0$  and all  $n \in \mathbb{N}$ . By Lemma 2.3(i), there exist a subsequence  $(\mu_{n_k})$  of  $(\mu_n)$ , a sequence  $(X_k)$  in  $Y^\Phi$  and  $X \in Y^\Phi$  such that  $X_k$  has distribution  $\mu_{n_k}$  for each  $k \in \mathbb{N}$ ,  $X$  has distribution  $\mu_0$ , and  $X_k \xrightarrow{o} X$  in  $Y^\Phi$ . Then (ii) implies that

$$\mathcal{R}_\rho(\mu_{n_k}) = \rho(X_k) \rightarrow \rho(X) = \mathcal{R}_\rho(\mu_0).$$

This contradicts (2.4) and proves (ii)  $\implies$  (i). The reverse implication (i)  $\implies$  (ii) is immediate by Lemma 2.3(ii). □

We now present the proof of Theorem 2.2.

*Proof of Theorem 2.2.* Suppose that  $\rho: L^\Phi \rightarrow \mathbb{R}$  is law-invariant and order-continuous. Take any  $X \in L^\Phi$  and any stationary ergodic sequence of random variables with the same distribution as  $X$ . Denote by  $\mu_0$  their distribution. Let  $\lambda > 0$  be such that  $\mathbb{E}[\Phi(\lambda|X|)] < \infty$ . Put  $\Phi_\lambda(\cdot) := \Phi(\lambda \cdot)$ . Arguing similarly as in the proof of Krättschmer *et al.* (2014, Theorem 2.6), by applying Birkhoff's ergodic theorem, one obtains a measurable subset  $\Omega_0$  of  $\Omega$  such that  $\mathbb{P}(\Omega_0) = 1$  and for every  $\omega \in \Omega_0$ ,

$$\widehat{m}_n(\omega) \xrightarrow{\Phi_\lambda\text{-weakly}} \mu_0.$$

Since  $\rho$  is order-continuous on  $L^\Phi$  and  $Y^{\Phi_\lambda} \subset L^\Phi$ ,  $\rho$  is also order-continuous on  $Y^{\Phi_\lambda}$ . By Lemma 2.4,  $\mathcal{R}_\rho$  is continuous on  $\mathcal{M}(Y^{\Phi_\lambda})$  with the  $\Phi_\lambda$ -weak topology. Thus,  $\widehat{\rho}_n(\omega) = \mathcal{R}_\rho(\widehat{m}_n(\omega)) \rightarrow \mathcal{R}_\rho(\mu_0) = \rho(X)$  for every  $\omega \in \Omega_0$ . This proves that  $\rho$  is strongly consistent at  $X$ .  $\square$

**Remark 2.5.** In our definition of Orlicz spaces, we do not allow the Orlicz function to take the  $\infty$  value, which excludes  $L^\infty$  from the above considerations. However, Theorem 2.2 remains true for  $L^\infty$ . Let  $\rho: L^\infty \rightarrow \mathbb{R}$  be an order-continuous, law-invariant functional. Take any  $X \in L^\infty$  and any stationary ergodic sequence  $(X_n)$  of random variables with the same distribution as  $X$ . Denote by  $\mu_0$  their distribution. By Birkhoff's ergodic theorem and an application of Theorem 6.6 in Parthasarathy (1967, Chapter 1), there exists a measurable subset  $\Omega_0$  of  $\Omega$  such that  $\mathbb{P}(\Omega_0) = 1$  and  $\widehat{m}_n(\omega) \xrightarrow{\text{weakly}} \mu_0$  for any  $\omega \in \Omega_0$ . One may assume further that  $|X(\omega)| \leq \|X\|_\infty$  and  $|X_n(\omega)| \leq \|X\|_\infty$  for any  $\omega \in \Omega_0$  and any  $n \geq 1$ . Fix any  $\omega \in \Omega_0$ . The classical Skorohod representation yields  $(X_n^\omega) \subset L^\infty$  and  $X^\omega \in L^\infty$  such that  $X_n^\omega \sim \widehat{m}_n(\omega)$  for every  $n \in \mathbb{N}$ ,  $X^\omega \sim \mu_0$ , and  $X_n^\omega \xrightarrow{a.s.} X^\omega$ . We may assume that  $|X_n^\omega| \leq \|X\|_\infty$  on  $\Omega$  for every  $n \geq 1$ . It follows that  $X_n^\omega \xrightarrow{o} X^\omega$  in  $L^\infty$ . Hence, by order continuity of  $\rho$ , we get  $\widehat{\rho}_n(\omega) = \mathcal{R}_\rho(\widehat{m}_n(\omega)) = \rho(X_n^\omega) \rightarrow \rho(X^\omega) = \rho(X)$ . This proves that  $\rho$  is strongly consistent on  $L^\infty$ .

Order continuity of law-invariant functionals on Orlicz spaces is generally stronger than norm continuity. In the following, we show that it is satisfied by a large class of risk measures, namely, spectral risk measures. Spectral risk measures were introduced in Acerbi (2002) and includes many important risk measures such as the expected shortfall. Let  $\phi$  be a nonnegative and nondecreasing function such that  $\int_0^1 \phi(t)dt = 1$  ( $\phi$  is called a spectrum). The associated spectral risk measure is defined by:

$$\rho_\phi(X) = \int_0^1 \phi(t)F_X^{-1}(t) dt, \quad X \in L^1$$

where  $F_X^{-1}(t) = \inf\{x \in \mathbb{R}: F(x) \geq t\}$  is the left quantile function of  $X$ . It is known that  $\rho_\phi$  takes values in  $(-\infty, \infty]$  and is convex, monotone, and lower semicontinuous with respect to the  $L^1$  norm (see e.g., Amarante and Liebrich 2024, Lemma C.1). For spectral risk measures, the empirical estimator  $\widehat{\rho}_n$  has the form of an  $L$ -statistic and the strong consistency can be studied using tools from the theory of  $L$ -statistics (see e.g., Tsukahara 2014). Below we give a simple proof of the strong consistency of  $\rho_\phi$  in the Orlicz space framework based on Theorem 2.2.

**Corollary 2.6.** Let  $\phi$  be a spectrum function such that  $\rho_\phi$  is real-valued on  $L^\Phi$ . Then  $\rho_\phi$  is order continuous on  $L^\Phi$  and is thus strongly consistent everywhere on  $L^\Phi$ .

*Proof.* Suppose that  $X_n \xrightarrow{a.s.} X$  and there exists  $Y \in L^\Phi$  such that  $|X_n| \leq Y$  for every  $n \in \mathbb{N}$ . It is well known that  $F_{X_n}^{-1} \xrightarrow{a.s.} F_X^{-1}$  on  $(0,1)$  (with the Lebesgue measure). Hence,

$$\phi F_{X_n}^{-1} \xrightarrow{a.s.} \phi F_X^{-1}.$$

Next, note that since  $-Y \leq X_n \leq Y$ ,  $F_{-Y}^{-1} \leq F_{X_n}^{-1} \leq F_Y^{-1}$  on  $(0,1)$ . Thus,

$$\phi F_{-Y}^{-1} \leq \phi F_{X_n}^{-1} \leq \phi F_Y^{-1} \quad \text{for every } n \in \mathbb{N}.$$

Since  $\rho$  is real-valued on  $L^\Phi$ ,  $\phi F_Y^{-1} \in L^1$  and  $\phi F_{-Y}^{-1} \in L^1$ . Thus, by the dominated convergence theorem,

$$\rho_\phi(X_n) = \int_0^1 \phi(t) F_{X_n}^{-1}(t) dt \rightarrow \int_0^1 \phi(t) F_X^{-1}(t) dt = \rho_\phi(X).$$

This proves that  $\rho_\phi$  is order continuous on  $L^\Phi$ . The strong consistency follows from Theorem 2.2.  $\square$

We end this note with the following remark that improves the implication (b)  $\implies$  (a) in Krätschmer *et al.* (2014, Theorem 2.8) due to our Theorem 1.4.

**Corollary 2.7.** *Suppose that  $\Phi$  satisfies the  $\Delta_2$ -condition. Let  $\rho$  be any convex, law-invariant, order-bounded-above functional on  $L^\Phi$ . Then  $\mathcal{R}_\rho$  is continuous on  $\mathcal{M}(L^\Phi)$  for the  $\Phi$ -weak topology.*

*Proof.* By Theorem 1.4,  $\rho$  is norm-continuous. When  $\Phi$  satisfies the  $\Delta_2$ -condition, order convergence in  $L^\Phi$  implies norm convergence. Thus,  $\rho$  is also order-continuous. Under the  $\Delta_2$ -condition, we also have  $H^\Phi = Y^\Phi = L^\Phi$ . Now apply Lemma 2.4.  $\square$

**Acknowledgement.** The authors acknowledge the support of NSERC Discovery Grants.

## References

- Acerbi, C. (2002) Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**(7), 1505–1518.
- Aliprantis, C.D. and Burkinshaw, O. (1998) *Principles of Real Analysis*. California: Academic Press.
- Aliprantis, C.D. and Burkinshaw, O. (2006) *Positive Operators*. Dordrecht: Springer.
- Amarante, M. and Liebrich, F.B. (2024) Distortion risk measures: Prudence, coherence, and the expected shortfall. *Mathematical Finance*, **34**(4), 1291–1327.
- Artzner, Ph., Delbaen, F., Eber, J.-M. and Heath, D. (1999) Coherent measures of risk. *Mathematical Finance*, **9**, 203–228.
- Bellini, F., Fadina, T., Wang, R. and Wei, Y. (2022) Parametric measures of variability induced by risk measures. *Insurance, Mathematics and Economics*, **106**, 270–284.
- Bellini, F., Laeven, R.J.A. and Rosazza Gianin, E. (2021) Dynamic robust Orlicz premia and Haezendonck–Goovaerts risk measures. *European Journal of Operational Research*, **291**(2), 438–446.
- Bellini, F. and Rosazza Gianin, E. (2012) Haezendonck–Goovaerts risk measures and Orlicz quantiles. *Insurance: Mathematics and Economics*, **51**, 107–114.
- Biagini, S. and Frittelli, M. (2009) On the extension of the Namioka–Klee theorem and on the Fatou Property for risk measures. In *Optimality and Risk: Modern Trends in Mathematical Finance* (eds. F. Delbaen, M. Rasonyi and C. Stricker), pp. 1–28. Berlin: Springer.
- Biagini, S., Frittelli, M. and Grasselli M. (2011) Indifference price with general semimartingales. *Mathematical Finance*, **21**(3), 423–446.
- Breiman, L. (1991) *Probability*. Classics in Applied Mathematics, vol. 7. Philadelphia: SIAM (corrected reprint of the 1968 original).
- Chen, S., Gao, N., Leung, D. and Li, L. (2022) Automatic fatou property of law-invariant risk measures. *Insurance: Mathematics and Economics*, **105**, 41–53.
- Edgar, G.A. and Sucheston, L. (1992) *Stopping Times and Directed Processes*. Cambridge: Cambridge University Press.
- Farkas, W., Koch-Medina, P. and Munari, C. (2014) Beyond cash-additive risk measures: When changing the numéraire fails. *Finance and Stochastics*, **18**, 145–173.
- Föllmer, H. and Schied, A. (2011) *Stochastic Finance. An Introduction in Discrete Time*, 3rd edn. Berlin: de Gruyter.
- Gao, N., Leung, D., Munari, C. and Xanthos, F. (2018) Fatou property, representations, and extensions of law-invariant risk measures on general Orlicz spaces. *Finance and Stochastics*, **22**, 395–415.
- Gao, N., Leung, D. and Xanthos, F. (2019). Closedness of convex sets in Orlicz spaces with applications to dual representation of risk measures. *Studia Mathematica*, **249**, 329–347.
- Gao, N., Munari, C. and Xanthos, F. (2020) Stability properties of Haezendonck–Goovaerts premium principles. *Insurance: Mathematics and Economics*, **94**, 94–99.
- Gao, N. and Xanthos, F. (2018) On the C-property and  $w^*$ -representations of risk measures. *Mathematical Finance*, **28**(2), 748–754.
- Jouini, E., Schachermayer, W. and Touzi, N. (2006) Law invariant risk measures have the Fatou Property. In *Advances in Mathematical Economics*, vol. **9**, pp 49–71.
- Krätschmer, V., Schied, A. and Zähle, H. (2014) Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.
- Krätschmer, V., Schied, A. and Zähle, H. (2017) Domains of weak continuity of statistical functionals with a view toward robust statistics. *Journal of Multivariate Analysis*, **158**, 1–19.

- Parthasarathy, K.R. (1967) *Probability Measures on Metric Spaces*. *Probability and Mathematical Statistics*, vol. **3**. New York: Academic Press.
- Rockafellar, R.T., Uryasev, S. and Zabarankin, M. (2006) Generalized deviation in risk analysis. *Finance and Stochastics*, **10**, 51–74.
- Ruszczynski, A. and Shapiro, A. (2006). Optimization of convex risk functions. *Mathematics of Operations Research*, **31**(3), 433–452.
- Shapiro, A. (2013) Consistency of sample estimates of risk averse stochastic programs. *Journal of Applied Probability*, **50**(2), 533–541.
- Tsukahara, H. (2014) Estimation of distortion risk measures. *Journal of Financial Econometrics*, **12**(1), 213–235.
- Villani, C. (2021) *Topics in Optimal Transportation*, vol. **58**. American Mathematical Society.